Statistical Analysis of Kemp's Ridley Nesting Trends

Philip M. Dixon and Selina S. Heppell

FINAL

30 August 2015

Table of Contents

Figures

Tables

Abstract

To examine potential changes in nesting trends and expected nest counts for Kemp's ridley sea turtle following the observed reduction in nests in 2010, we build regression and state-space time series models for the number of Kemp's ridley nests from the early or mid-1990s through 2009. These models are fit to nest counts from seven individual nesting beach areas and three beach sums. The posterior distribution of parameter estimates is used to estimate the projected number of nests in 2010, 2011, 2012, 2013, and 2014 and to estimate the difference between projected number of nests and the observed number of nests. The average rate of increase prior to 2010 for all nesting areas combined was 17% per year. For most nesting beach areas and all sums, observed nest counts were below the 95% confidence interval of expected nests in 2010, 2013 and 2014, and were below the median expected nests in 2011 and 2012. For the total nest count summed over all seven nesting beach areas, the median differences (projected - observed) are 10,730 nests, 8,110 nests, 11,990 nests, 23,480 nests, and 35,200 nests in 2010, 2011, 2012, 2013, and 2014 respectively. The loss in 2010 is 49% of the total number of nests in 2009. For the total nest count on beaches in Texas, the median differences are 110 nests, 120 nests, 220 nests, 410 nests, and 620 nests in 2010, 2011, 2012, 2013 and 2014. The loss in 2010 is 56% of the total number of nests in Texas in 2009. Correlation analyses indicate a high degree of correlation among nesting sites over time, but relatively low correlation in the year-to-year change in nest counts once the trend is removed. Our analysis indicates a large reduction in the number of nests that would have been seen had the population continued on the trajectory that was established prior to 2009.

Introduction

The Kemp's ridley sea turtle, *Lepidochelys kempii*, is considered the world's most endangered sea turtle (Plotkin 2007). The species is genetically distinct from the similar olive ridley and loggerhead turtles (Bowen et al. 1991) and has a restricted distribution, nesting almost exclusively in the western Gulf of Mexico on sandy beaches from Vera Cruz, Mexico, to southern Texas, USA (Pritchard 2007). Abundant in the middle of the 20th century, with estimates of tens of thousands of nests laid on a single day (Hildebrand 1963), the species suffered near extinction due to harvest of eggs and adults as well as high incidental mortality in fishing gear (Márquez-M. et al. 1999; NRC 1990). In the 1980s, the species began to recover, thanks to a bi-national effort that included protection of nearly all nests laid each year and laws in both countries that reduced at sea mortality in shrimp nets and other fishing gear. Nest counts began to increase in 1985, and accelerated in the 1990s to a rate of over 15% per year on average (TEWG 2000). The recovery was considered a great success story, with over 21,000 nests counted in 2009, and potential downlisting from endangered to threatened status was expected when the annual nest count was due to hit 25,000 (NMFS/USFWS/SEMARNAT 2011; Crowder and Heppell 2011). But in 2010, the nest counts dropped by nearly 40%, the largest single year reduction recorded since intensive recovery efforts began in 1978. Nest counts returned to levels comparable to 2009 in 2011 and 2012, but dropped again in 2013 and 2014. The 2014 count for monitored beaches in Mexico and Texas was only a little over 12,000, the lowest count since 2006.

1

The apparent change in population trajectory of the Kemp's ridley is important to investigate because of the endangered status of the species and ongoing efforts to document impacts to wildlife that may be linked to the Deepwater Horizon (DWH) spill. Because all of the nests are counted and protected through most of the Kemp's ridley range (Heppell et al. 2004), we have an unprecedented opportunity to test for changes in a key index of abundance for an entire species over a long time period. Nest monitoring has occurred through much of Tamaulipas as the range of the species has expanded during the recovery. Nest counts have been conducted in Texas since the 1990s, where nests had steadily increased every year until 2010 (Shaver 2005). Nests are counted daily and most are moved to protected areas ("corrals") so eggs can be recorded, incubated and released as hatchlings. Data are available through the National Park Service (Texas) and a US-Mexico funded project run by CONANP and the Gladys Porter Zoo (Mexico). The number of years of data available for analysis varies by beach because new monitoring areas were established as the range of the species expanded over time (Heppell et al. 2004).

We used a series of analyses on nest counts from each nesting beach and pooled data to test the following questions:

1) Is the observed number of nests in 2010-2014 different from a predicted count, given the observed fluctuations in nest counts in previous years?
2) If the observed nest counts are different from expected, what is the deviation in potential nests?
3) Are nest counts correlated across nesting areas in Texas and Mexico?

Our model predicts the nest count for each year and estimates the deviations from the predicted nest count in the five post-spill years (2010, 2011, 2012, 2013, and 2014). Comparison of the observed nest count in 2010 or later to the predicted nest count indicates whether something unusual happened in that year. The analysis is based on time-series intervention analysis (Box and Tiao 1975) using models that do not pre-specify the form of the intervention (Harvey 1989, p. 402 – 404). We develop a regression model based on biological information and data characteristics. Because the data are collected over time, we assess variations on the regression model that include autocorrelated residuals.

We predict the distribution of the difference between observed and expected counts in the five post-spill years using a Bayesian state-space time series implementation of the regression model (Harvey 1993). This model is fit to the data up to 2009, then samples from the joint posterior distribution of parameters are used to predict the nest count in 2010, 2011, and beyond. These predictions require both the predicted mean nest count in a future year (e.g. 2010) and the predicted variance of that count. These predictions assume that the time series structure observed from the start of the data set through 2009 continue forward in time.

We analyze the correlations between nests counts in Texas and each Mexican nesting area. We evaluate both the correlation between contemporaneous nest counts and the correlation between annual changes in nest counts.

2

Methods

Nesting Data

The data are the time series of annual counts of the number of Kemp's ridley turtle nests on individual nesting beaches or beach aggregates in northern Mexico and Texas (Table 1), provided by USFWS and the Texas Parks and Wildlife Department. Figure 1 shows the locations of these nesting areas. Nesting in Texas is spread over many beaches, with few nests per year per beach. Mexican nesting areas may also include multiple beaches, and are defined by the name and location of monitoring camps (Heppell et al. 2004). For these analyses, we sum the annual nest count for all Texas beaches and use the nesting beach camp names provided in binational reports to USFWS that are submitted annually by the Gladys Porter Zoo, Brownsville, Texas.

The initial year of our analysis varies between 1991 for Tepehuajes, Rancho Nuevo, and Barra del Tordo in Mexico and 1996 or 1997 for other sites in Mexico and Texas (Table 1, Figure 2). Although some data for three Mexican sites (Tepehuajes, Rancho Nuevo, and Barra del Tordo) are available before 1991, we used only data from 1991 onwards because some beaches were not monitored prior to 1991. Tepehuajes, Rancho Nuevo, and Barra del Tordo have been traditionally used as the primary index of abundance for the species over the monitoring time period, as it was agreed by the Turtle Expert Working Group (1998) that these beaches represented the core nesting area of the species. Thus, our analyses were conducted on the nest counts for these three beach areas in aggregate (Sum3) as well as for the individual areas and a pool of all counts. We include data through the 2014 nesting season in the analyses.

SumMX is the sum of all six Mexican nesting sites specified in the reports, and SumAll is the sum of all Texas and Mexico nesting sites. The annual nest counts analyzed here are provided in Appendix 1.

Table 1. Nesting areas used in these analyses, with length of data record and data source.

Nesting Beach (North to South)	Years of Data	Aggregates that Include these Data	Data Source
Texas beaches[a]	1997-2014	SumAll	Dr. Donna Shaver, National Park Service
La Pesca	1996-2014	SumMX, SumAll	Gladys Porter Zoo reports[b]
Tepehuajes	1991-2014	Sum3, SumMX, SumAll	
Rancho Nuevo	1991-2014	Sum3, SumMX, SumAll	
Barra del Tordo	1991-2014	Sum3, SumMX, SumAll	
Altamira & Madero	1996-2014	SumMX, SumAll	
Miramar	1997-2014	SumMX, SumAll	
Notes			

Notes
   a. These include: Bolivar Peninsula, Galveston Island, Brazoria County, N. of Surfside, Surfside Beach, Quintana Beach, Bryan Beach, Brazoria County, N. of Sargent Beach, Sargent Beach, Matagorda Peninsula, Matagorda Island, San Jose Island, Mustang Island, Corpus Christi Bay, North Padre Island, North of PAIS, Padre Island National Seashore (PAIS), South Padre Island, and Boca Chica Beach.
   b. USFWS 1997-2014, "Final Report on the Mexico / United States of America Population Restoration Project for the Kemp's Ridley Sea Turtle, *Lepidochelys kempii,* on the Coasts of Tamaulipas, Mexico." Cooperative

Figure 1. Locations of study beaches on the Gulf coast of northern Mexico and Texas
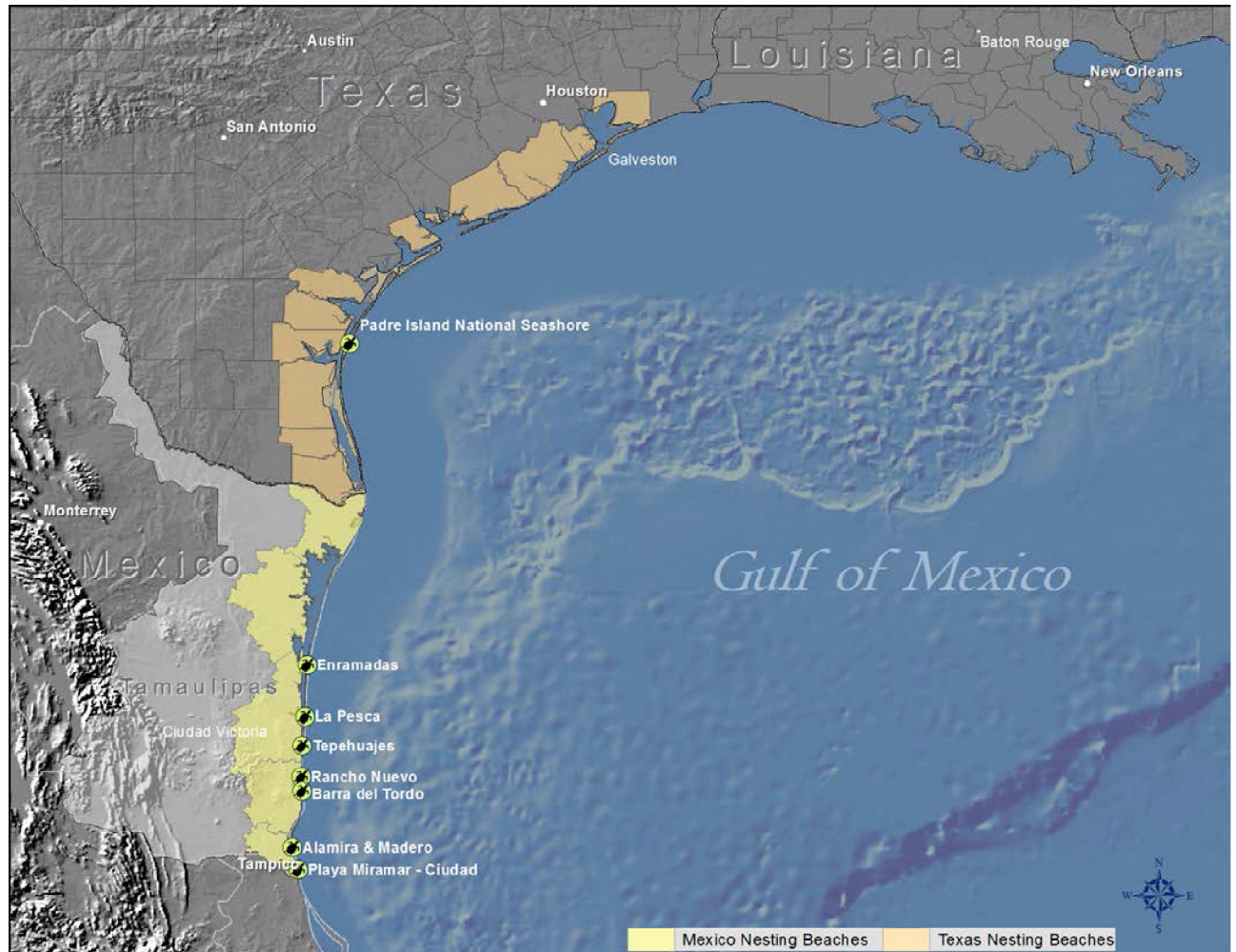
Figure 2. Nest count data by nesting area. A: 1991-2009, B: 2008-2014. Rancho Nuevo, Mexico counts are scaled on the second access. Data provided by Mexico / United States of America Population Restoration Project for the Kemp's Ridley Sea Turtle, *Lepidochelys kempii,* on the Coasts of Tamaulipas, Mexico." Cooperative Grant 201814J833 (annual reports). Raw data in Appendix 1. Data provided for Texas by National Park Service.

A. 1991-2009
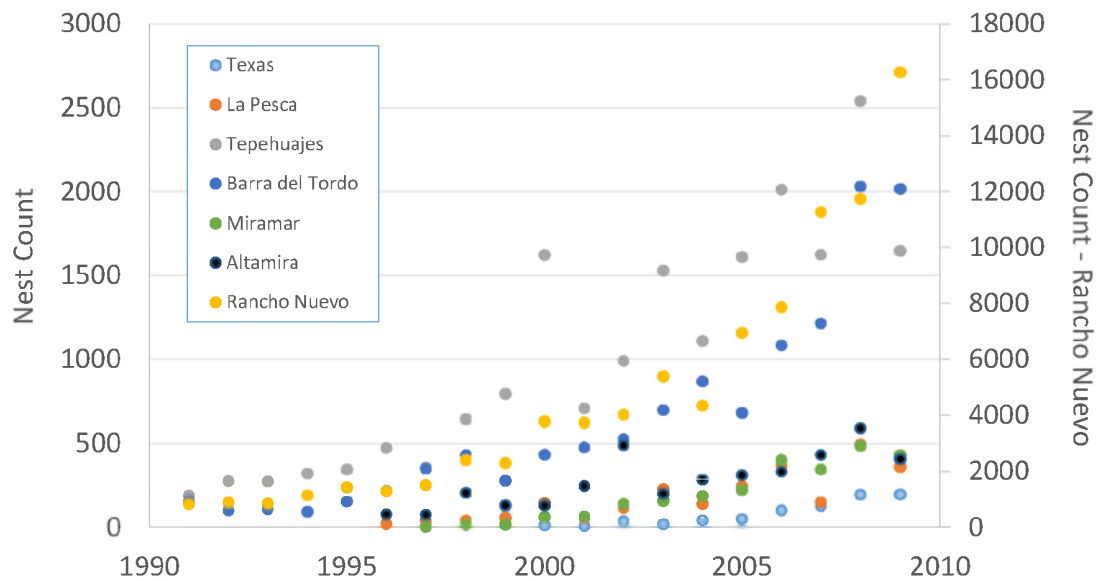
B. 2008-2014
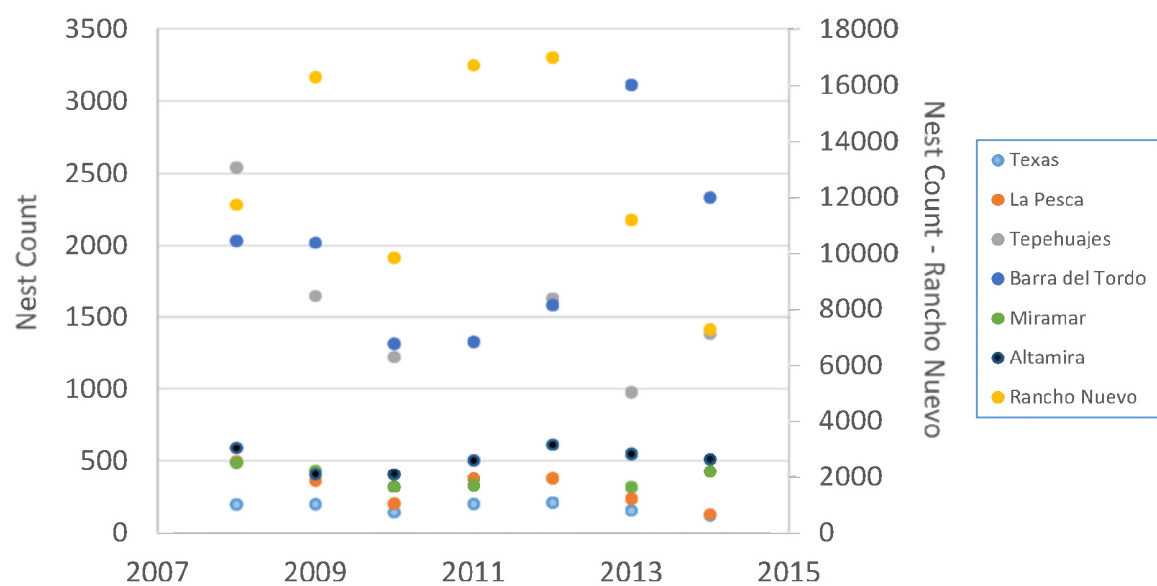
Preliminary Inspection of the Data

We evaluated characteristics of the data so we can formulate and justify a statistical model for how nest counts vary over time. This preliminary inspection considers the nature of the trend over time, the distribution of observations around that trend, and the statistical characteristics of the deviations of observations around the trend.

The preliminary inspection evaluates two counts: the Sum3 count (the sum of annual nests at Tepehuajes, Rancho Nuevo, and Barra del Tordo sites in Mexico) and the SumAll count (the sum of nest counts for all Texas and Mexico nesting beaches). The Sum3 count is available since 1991, longer than any other beach or beach sum, and includes from 87% to 100% of the annual total number of nests for the species (Márquez-M. et al. 2001; Heppell et al. 2007). These 3 nesting areas were the only known source of nests prior to the mid 1990s (Turtle Expert Working Group 1998), and continue to represent over 90% of nests counted today. As the species has expanded north and south of these core areas, the proportion of nests counted at Rancho Nuevo, Tepehuajes and Barra del Tordo has decreased gradually, but the annual change in nest counts has not been consistently higher or lower than the change observed across all nest counts.

SumAll was chosen for variance evaluation because that count is the focus of the statistical modeling. Data are available from 1997, the first year that data are available from all of the nesting beaches. Similar patterns (data not shown) are found in SumMX, the sum of all beaches in Mexico and most of the individual beaches. The preliminary inspection of the data considers only the nest counts through 2009, and is detailed in Appendix 2.

For both the Sum3 and SumAll counts, the trend in the observed numbers of nests is roughly exponential, although there is some additional variation from year to year (Figure 3). The trend is approximately linear when the observed numbers of nests are plotted on a log scale (Figure 4). A variety of different models were examined for trend evaluation; best fits were obtained for those with a log linear trend, and there was no evidence for a trend in variance (Appendix 2).

6

Figure 3. Annual number of nests for Tepehuajes, Rancho Nuevo, and Barra del Tordo beaches summed (blue) and for all nesting beaches summed (red).
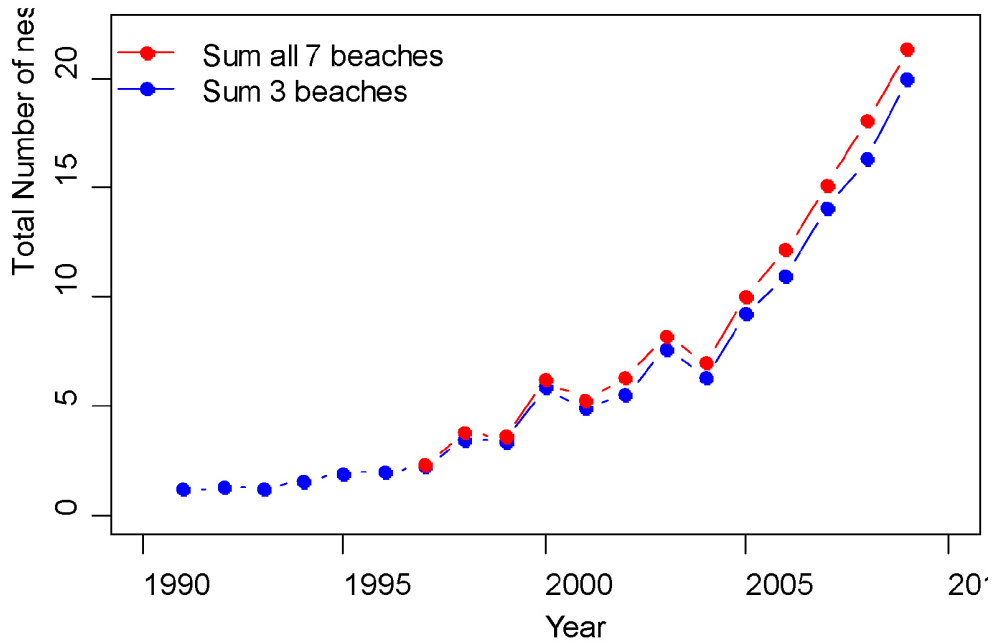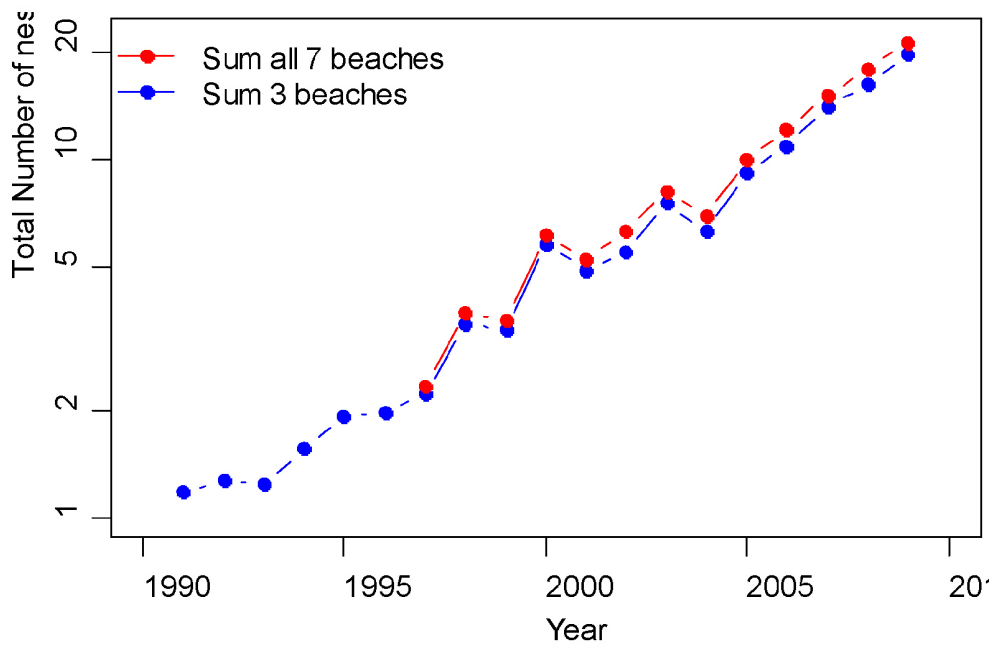

Figure 4. Annual number of nests for Tepehuajes, Rancho Nuevo, and Barra del Tordo beaches summed (blue) and for all nesting beaches summed (red), plotted on a log scale for the Y axis.

7

Statistical Models

The models used for the analysis are motivated by biological considerations and the features seen in the preliminary inspection of the data. The biological considerations for Kemp's ridley turtles include:

- The nest counts for each year and each nesting area are, with minor exceptions, known.
- These counts depend on the number of nesting females, which is an unknown proportion of the number of mature females that likely varies through time because females do not typically nest annually (Heppell et al. 2004; NMFS/USFWS/SEMARNAT 2011 2011).
- Other population modeling (Heppell et al. 2004; NMFS/USFWS/SEMARNAT 2011 2011) fit to nest counts back to 1978 suggests that the population of mature adults has been increasing exponentially since the early 1990s.

The features of the Sum3 and SumAll data that inform potential models are:

- The nest count is increasing approximately exponentially, with some variation from year to year.
- Annual nest counts are overdispersed, with more variability than expected under a Poisson model (Appendix 2).
- The variability in log transformed nest count around a simple linear regression line has approximately constant variance.
- Residuals from a log linear regression show no evidence of autocorrelation at lags of 1, 2, or 3 years. Model comparison shows that a model with uncorrelated errors is more appropriate than models with AR(1) or MA(1) correlation structures (Appendix 2).

Regression Models

We use linear regression to fit an intervention model to the data through 2014. This model is separately fit to each beach or beach sum, except Miramar, for which the linear model was not appropriate (Table S1). We allow the number of nests in each year after 2009 to deviate from the trend up to 2009 and make no assumptions about the form or intensity of those deviations. This is done by including year-specific deviations for 2010, 2011, 2012, 2013 and 2014 (Equation 1).

$$\log N_t = \beta_0 + \beta_1\, t + \beta_{2010}\, I_{2010} + \beta_{2011}\, I_{2011} + \beta_{2012}\, I_{2012} + \beta_{2013}\, I_{2013} + \beta_{2014}\, I_{2014} + \varepsilon_t$$

$$\tag{1}$$

$$\varepsilon_t \sim Normal(0,\ \sigma^2)$$

I_{2010} is an indicator variable that has the value of 1 in the year 2010 and the value of 0 in all other years. The other four indicator variables are defined similarly. The regression coefficients associated with each indicator variable quantify the deviation in log transformed nest count for that year.

Converting the results from fitting Equation (1) into an estimate of the difference between observed and expected nests in 2010-2014 requires predicting the distribution of the expected number of nests for each year after 2009. This computation is more easily done using a Bayesian framework. In this framework, the joint posterior distribution of all model parameters, including the variance of observations around the regression line, is estimated using the nest counts up to 2009. The posterior

8

distribution of the number of nests in 2010 or any subsequent year is then estimated from the joint posterior distribution of model parameters. Because the nest counts on Miramar show substantial lack of fit to a log-linear trend, we used state-space time series models to implement both the linear regression model and a more flexible model where needed, as for Miramar.

State-space Time Series Models

State-space time series models (Harvey 1993, p. 82-85), also called structural time series models (Harvey 1989, p. 10), or dynamic linear models (West and Harrison 1997, p. 32-34 provide a flexible set of models for the log transformed nest counts, $\log N_t$, over time. A simple state-space model is equivalent to a linear regression model with uncorrelated errors. Extensions of that simple model provide more flexible models for the trend over time and simple ways to model the autocorrelation between observations over time. State-space models have two components: an observation model and a process model (Harvey 1993, p. 82). The observation model connects the observed nest counts to an unobserved and unmeasured latent quantity, which in this problem can be considered to be the number of mature females in the population. The process model describes how that latent quantity changes over time. Different assumptions about how the number of mature females changes over time lead to different choices of the process model. Although using a model with two components to describe one time series may seem an unnecessary complication, the use of two components leads to parameters that are more interpretable than the parameters in other models for autocorrelated time-series data (Harvey 1993, p. 106).

Based on the preliminary inspection of the data, a reasonable observation model is

$$\log N_t = Z_t + \varepsilon_t. \tag{2}$$

$$\varepsilon_t \sim N(0, \sigma_{obs}^2)$$

The log transformed number of nests for a beach and year is modeled as a random deviation from a latent log-scale number of mature females, Z_t. The random deviation, ε_t, accounts for the unpredictable year-to-year variation in the fraction of mature females that nest and the number of nests laid per female.

Different assumptions about the change in the number of mature females lead to different choices of the process model. Consequently, we considered three process models. The first ("constant trend") model is equivalent to a linear regression (Equation S2). In this model, the mean year-to-year change in log transformed nest count is constant. This model can be written in state-space form as the process model in Equation (3) used with the observation model of Equation (2).

$$Z_i = Z_{i-1} + b \tag{3}$$

The Z_t terms in Equation (3) are the latent, unobserved, log-scale number of mature females; the b is the constant annual increase (trend). The second ("process error") model generalizes Equation (3) to allow random year-to-year fluctuation in the latent variable, Z_t. In this model, the year-to-year change in

9

mean log transformed nest count is not a constant. Instead, the year-to-year change includes a random effect (process error) for each year.

$$Z_t = Z_{t-1} + b + \tau_t \tag{4}$$

$$\tau_t \sim N(0, \sigma_{process}^2)$$

Unlike ε_t, the error in the observation equation (Equation 2) that only influences the observation for year t, the process error, τ_t, influences the mean count for that year and all subsequent years. Two possible mechanisms that would generate process errors are year-to-year variation in survival and nesting probability of mature adults or variation in year-class strength for recruits to the nesting population.

The third ("varying trend") model generalizes Equation (4) to allow the mean trend (year-to-year change) to vary smoothly over time.

$$Z_t = Z_{t-1} + b_t + \tau_t \tag{5}$$

$$\tau_t \sim N(0, \sigma_{process}^2)$$

$$b_t = b_{t-1} + \omega_t \tag{6}$$

$$\omega_t \sim N(0, \sigma_{trend}^2)$$

Equation (6) allows the mean trend, b_t, to vary over time. One possible mechanism for which such a model would be appropriate is a situation where the population exponential growth rate is increasing (or decreasing) over time.

The three variance parameters, σ_{obs}^2, $\sigma_{process}^2$, and σ_{trend}^2, quantify the year-year idiosyncratic variability of the observed nest count, the process error in the latent population size, and the variability in the exponential growth rate parameter. The third (varying trend) model is the most general; the first and second models are simplifications of the third model with one or two variance parameters set to 0. The second (process error) model is the third (varying trend) model with $\sigma_{trend}^2 = 0$. The first (constant trend) model is the third model with $\sigma_{trend}^2 = 0$ and $\sigma_{process}^2 = 0$.

Both the process error and varying trend models allow autocorrelation between observations in adjacent years. The nature and type of that autocorrelation depend on the model and the relative magnitudes of the variance parameters, σ_{obs}^2, $\sigma_{process}^2$, and σ_{trend}^2 (Harvey 1993, p. 120-123).

The constant trend model (Equations 2 and 3) was fit to the nest counts for six beaches and all three beach sums. This model is the state-space equivalent of the linear regression model with uncorrelated errors. The varying trend model was fit to the nest counts from Miramar beach because those data show substantial lack of fit to a linear model (see Appendix 2).

The state-space models are standard time-series models that helped us explore a set of viable alternatives to explain observed variance in nest counts and identify the best fitting, most parsimonious

10

model. The simplest is the exponential growth model with a constant unobserved population growth rate and a constant proportion nesting. The process error model allows the unobserved population growth rate to vary erratically from year to year for unknown reasons. The varying trend model allows the mean population growth rate to change systematically over time, without specifying the form of that systematic change.

Model parameters were estimated by Bayesian Markov Chain Monte Carlo (MCMC) methods (West and Harrison 1997). Diffuse proper priors were used for the initial values of the latent population size, Z_0, and trend, b_0. Uniform distributions for the standard deviations were used as the prior distributions for the three variance components, as recommended by Gelman (2006). Computing was done using R and the BRugs library with the R2WinBUGS interface. Randomly chosen starting values were used to initialize three independent chains. The first 10,000 samples from each chain were discarded as burn-in. The subsequent 30,000 samples were thinned 30-fold to give 3,000 samples from the joint posterior distribution of all parameters. Convergence was assessed using the Brooks-Gelman-Rubin statistic (West and Harrison 1997), which was close to 1 in all cases.

In all cases, model parameters were estimated using only data up to and including 2009. We then projected the model forward in time to estimate the distributions of nest counts in 2010 through 2014 under the assumption that the dynamics in 2010 through 2014 were a continuation of the dynamics through 2009. The posterior distributions of nest counts for each site in 2010 through 2014 was simulated by drawing a sample from the posterior distributions of all parameters (e.g., Z_{2009}, b and σ^2_{obs} for the constant trend model), simulating realizations of ε_t (and τ_t and ω_t where appropriate) for each year (2010 through 2014) and computing $\log N_{2010}$, $\log N_{2011}$, $\log N_{2012}$, $\log N_{2013}$, and $\log N_{2014}$. This was repeated to generate 3,000 samples of the posterior distribution of nest counts in each of the five years following 2009. The posterior distribution was computed as the difference between the projected number of nests and observed nest counts. Naturally, the accuracy of this value as an indicator of "lost nests" may be reduced over time because the analysis assumes projection of the observed trend for up to 5 years beyond our observations. However, with nest counts alone, this simple projection is the only analysis possible or necessary to address the question of how nest numbers have changed over time.

We focus on results from the total number of nests for all nesting beaches (SumAll). Results for individual beaches and the other two beach sums are given in the supplemental material.

Correlations Between Beaches

We calculated the correlation between the nest count for Texas and the nest count in the same year for each Mexican nesting area. To smooth out some of the irregular fluctuations in the nest count, we also calculated the two-year running average nest count for each beach, and then calculated the correlations between Texas and each Mexican nesting area. These nest count correlations are all large because all populations are increasing through 2009. We also evaluated the correlations between annual fluctuations in the nest count by calculating the difference in log count between subsequent years for each beach, then calculating the correlations between Texas and each beach. We considered both

11

fluctuations in the nest count and fluctuations in the two-year running average. All calculations were done twice: once for the entire data record (through 2014) and once for the data through 2009.

R and BUGS code for all analyses are in Appendix 5.

Results

Modeling

For the SumAll nest count, the estimated trend (i.e. the coefficient for *t* in Equation 4) is 0.166 (se = 0.010). Up through 2009, the number of nests was increasing an average of 18% (= exp(0.166)) per year. The nest counts in each year after 2009 are lower than expected based on this trend (Table 2). For example, in 2010, the estimated regression coefficient is -0.590 (se = 0.164, p = 0.0041). This means the observed number of nests was 55% (computed as exp(-0.590), 95% confidence interval: 40%, 76%) of the expected number based on the trend seen through 2009. Results for individual nesting areas, Sum3 and SumMX are given in Appendix 3.

Table 2. Estimated regression coefficients for year-specific deviations from expected trend for the SumAll response, with their standard errors and p-values for the test of no deviation. The estimated regression coefficients for each indicator variable give the difference on a log scale between the projected number of nests (based on data through 2009) and what was observed in that year. These log-scale differences can be expressed as percentages of the expected nest count by exponentiating the estimate. The 95% confidence interval is calculated by exponentiating the 95% confidence interval for the regression estimate.

Year	Regression Coefficient Estimate (β_{year})	Std Error of β_{year}	p-value	Actual as Percent of Expected	95% Confidence Interval of Expected Value
2010	-0.590	0.163	0.0041	55%	(40%, 76%)
2011	-0.321	0.168	0.083	73%	(52%, 101%)
2012	-0.430	0.174	0.031	65%	(46%, 91%)
2013	-0.882	0.179	0.0005	41%	(29%, 59%)
2014	-1.355	0.186	< 0.0001	26%	(18%, 37%)

The state-space model was used to estimate the difference between the observed nest count and the expected nest count if the trend seen through 2009 continued through 2014. For the SumAll response, this was done using the constant trend model (Equations 2 and 3) because those data were appropriately fit by the linear regression model with uncorrelated errors. Comparison of the observed nest counts for 2010 through 2014 to the posterior prediction interval for each year (Figure 5) allows quantification of the difference. In 2010, 2013 and 2014, the observed nest count is substantially below the lower bound of the 95% prediction interval for that year's projected count. In 2012, the observed nest count is slightly below the lower bound. The difference between observed and predicted nest

counts can be expressed as either as a difference for an individual year or cumulative difference since 2009 (Table 3). Results for other beaches and beach sums are in Appendix 4.

Plots of the posterior predictive distributions for the 7 individual beaches and the other two beach sums are in the Supplemental Material, Figure S5.

Table 3. Estimated difference between observed nest counts and projected nest counts since 2010 for the sum of all nesting beaches, relative to the expected nest count if the trend to 2009 had continued (SumAll response). All numbers are rounded to the nearest 10. The median cumulative nest loss is not the sum of the median per year losses because per year losses are correlated.

	Nest Difference By Year		Cumulative Nest Difference	
Year	Median	95% Prediction Interval	Median	95% Prediction Interval
2010	10,730	(3070, 22270)	10,730	(3070, 22270)
2011	8,110	(-1190, 21440)	19,020	(5020, 38530)
2012	11,990	(660, 28530)	31,040	(9780, 62030)
2013	23,480	(9830, 45560)	55,390	(23430, 99510)
2014	35,200	(18620, 61330)	91,200	(47650, 151980)

Figure 5. Posterior predictive intervals for the total nest count (SumAll) in 2010 through 2014. Blue dots are the observed total number of nests from 2007 through 2014. Intervals are 95% posterior predictive intervals from the constant trend model. Dashed line shows the expected trend based on nest counts to 2009.
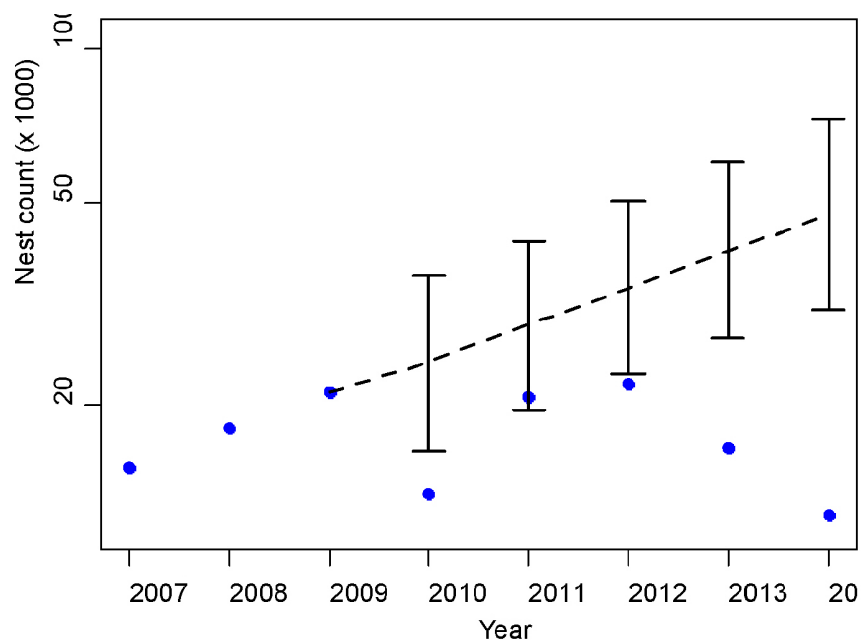
Correlations Between Nest Counts in Texas and Mexican Beaches

Figure 6 shows the correlations between nest counts in Texas and each beach in Mexico as a function of the latitude of the Mexican beach. Correlations for nest counts are above 0.8 for most beaches while correlations between the annual fluctuations are smaller. Correlations based on the two-year running average are similar to those based on the nest count, with one exception.

Figure 6. Correlations between Texas and each Mexican beach using the entire data record (through 2014).
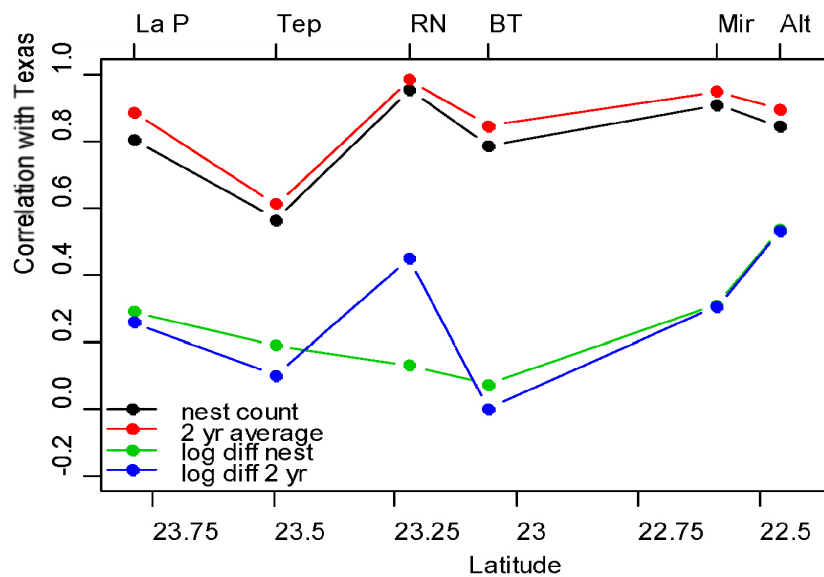
When correlations are calculated using only data through 2009 (Figure 7), those for nest counts are similar to those seen in the longer data record. Correlations in the annual fluctuations are smaller than those seen in the longer record and weakly negative for the annual fluctuation in the two-year running average.

Figure 7. Correlations between Texas and each Mexican beach using the data through 2009.
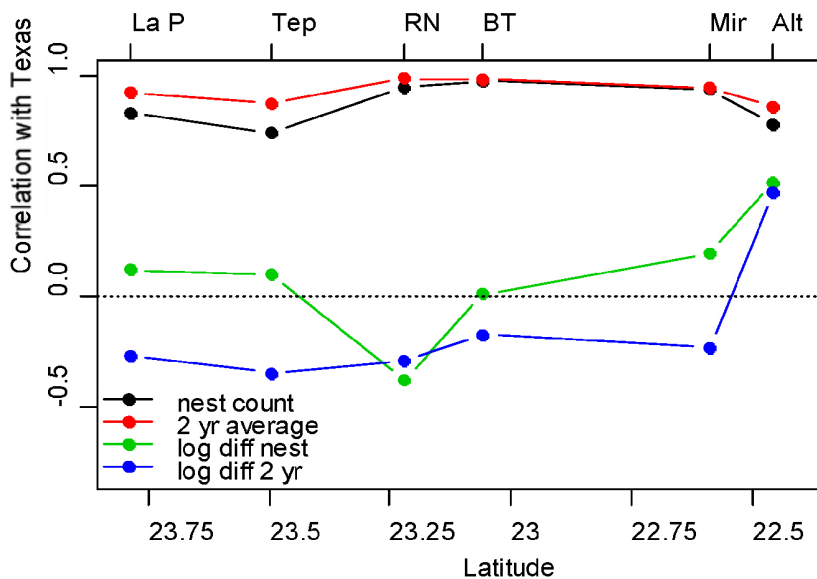
Discussion

All the models we explored indicate a clear deviation in expected nests based on the trend prior to 2010, even when year-to-year variance is considered. The largest deviations are in 2010, 2013 and 2014; nests were reduced relative to those expected in 2011 and 2012, but were not outside of the predicted intervals for all nesting areas. Despite the large drop in nest counts observed in 2010, the largest differences between observed and expected nests occur in 2013 and 2014, due to the model expectation of continued population growth. Nest "loss" from expected values is in the tens of thousands if we had seen the trend continue, but likely changes in this projection (due to mortality or other vital rates) cannot be determined from nest counts alone.

Despite variance in year to year counts at each site, overall, the nesting beach areas from Texas to southern Tamaulipas are well correlated prior to 2010. Continued monitoring at all sites is essential if there is an expectation of differential DWH impact in different areas.

We report the median difference in nest counts (projected – observed) rather than the mean difference because the median better represents the typical year. The projected nest difference has a right-skewed distribution so mean and the median are not the same. The mean would be more appropriate if we were considering many beaches or many years. However, our interest is in the nest difference for one beach or one beach sum in a specific year, which is more appropriately described by the median.

15

Use of the median is conservative; the median is smaller (fewer nests "lost") than is the mean for each year, because of the right-skewed distribution.

The 95% prediction intervals portray the uncertainty in the projected nest differences. The uncertainty in this difference is only that from the projection. The width of those intervals is strongly influenced by the year-to-year variability of nest counts around the fitted log-linear regression line. The width of the prediction intervals also depends on the model used to fit the data. Prediction intervals from the varying trend model are wider than those from the constant trend model because accounting for a varying trend requires additional sources of variability (the process error and the variability in the trend). The projected difference nests for Texas is relatively small, e.g. a median of 110 nests in 2010 and 120 nests in 2011. However, these losses represent large impacts because the nest counts observed in 2008 and 2009 are less than 200 (Appendix 1).

Our analysis indicates that a change occurred in 2010 that resulted in a large reduction in the expected number of nests, following a long-term, positive trend. Additional information on vital rate changes for juvenile and adult turtles throughout the Gulf of Mexico is needed to diagnose how and why nest numbers suddenly dropped and have thus far failed to recover. Importantly, the relationship between nest number and adult population size is not well known, and may not be constant through time.

Additional analysis and data needs include:

- Biological data to better understand process error in nest counts.
- Vital rate and recruitment information.
- More years of observation across the species' range

The scope of this analysis was limited to evaluation of nest counts alone. This analysis addressed the question of *how* things changed; determining *why* they changed requires hypothesis testing with an age structured model in addition to data on possible changes in reproductive rates and juvenile growth that could be contributing to the alteration of the nest count trajectory. While vital rates have not been monitored over time, there are other sources of information that could be used to improve our understanding of possible mechanisms for the change in nest counts. Future analysis, with more time, money and expertise, should be conducted by a Turtle Expert Working Group with multiple analysts to evaluate the situation more thoroughly

Acknowledgements

References

Bowen, B. W. 1991. Evolutionary distinctiveness of the endangered Kemp's ridley. *Nature* 352:22.

Box, G. E., and G.C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association, 70(349): 70-79.

Crowder, L.B. and S.S. Heppell. 2011. The decline and rise of a sea turtle: How Kemp's ridleys are recovering in the Gulf of Mexico. Solutions 2:67-73. http://www.thesolutionsjournal.com/node/859

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1: 515–533.

Harvey, A.C. 1993. *Time Series Models*, 2nd ed. MIT Press, Cambridge, MA.

Harvey, A.C. 1989. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, New York.

Heppell, S. S., D. Crouse, L. Crowder, S. Epperly, W. Gabriel, T. Henwood and R. Marquez. 2004. A population model to estimate recovery time, population size and management impacts on Kemp's ridley sea turtles. Chelonian Conservation and Biology 4:761-766.

Heppell, S. S., P. Burchfield, and J. Peña. 2007. Kemp's ridley recovery: How far have we come, and where are we headed? Pp. 325-335 in P. Plotkin, ed. Biology and Conservation of Ridley Sea Turtles. Johns Hopkins University Press, Baltimore.

Hildebrand, H. H. 1963. Halazgo del área de andación de la tortuga maria "lora" *Lepidochelys kempi* (Garman), en la costa occidental del Golfo de México. Sobretiro de Cienci, México 22:105-112.

Márquez-M., R., J. Díaz, M. Sánchez, P. Burchfield, A. Leo,, M. Carrasco, J. Peña, C. Jiménez, and R. Bravo. 1999. Results of the Kemp's ridley nesting beach conservation efforts in México. Marine Turtle Newsletter 85:2-4.

Márquez-M., R., P. Burchfield, M.A. Carrasco, C. Jiménez, J. Díaz, M. Garduño, A. Leo, J. Peña, R. Bravo and E. González. 2001. Update on the Kemp's ridley turtle nesting in México. Marine Turtle Newsletter 92: 2-4.

National Marine Fisheries Service, U.S. Fish and Wildlife Service, and SEMARNAT. 2011. Bi-National Recovery Plan for the Kemp's Ridley Sea Turtle (Lepidochelys kempii), Second Revision. National Marine Fisheries Service. Silver Spring, Maryland 156 pp. + appendices. http://www.fws.gov/kempsridley/Finals/kempsridley_revision2.pdf

National Research Council (NRC) 1990. *Decline of the Sea Turtle*. NRC Press, Washington, DC.

Plotkin, P.T. (ed) 2007. *Biology and Conservation of Ridley Turtles*. Johns Hopkins University Press, Baltimore.

Pritchard, P.C.H. 2007. Evolutionary relationships, osteology, morphology, and zoogeography of ridley sea turtles. Pp. 45-57 in P. Plotkin, ed. Biology and Conservation of Ridley Sea Turtles. Johns Hopkins University Press, Baltimore.

Shaver, D.J. 2005. Analysis of the Kemp's ridley imprinting and headstart project at Padre Island National Seashore, Texas, 1978–88, and subsequent Kemp's ridley nesting and stranding records on the Texas coast. Chelonian Conservation and Biology 4(4):846–859.

Turtle Expert Working Group. 1998. An assessment of the Kemp's ridley (*Lepidochelys kempii*) and loggerhead (*Caretta caretta*) sea turtle populations in the Western North Atlantic. NOAA Technical Memorandum NMFS-SEFSC-409. 96 pp.

Turtle Expert Working Group (TEWG). 2000. Assessment update for the Kemp's ridley and loggerhead sea turtle populations in the Western North Atlantic. U.S. Department of Commerce NOAA Technical Memorandum NMFS-SEFSC-444, 115 pp.

U.S. Fish and Wildlife Service, 1997-2014, "Final Report on the Mexico / United States of America Population Restoration Project for the Kemp's Ridley Sea Turtle, *Lepidochelys kempii*, on the Coasts of Tamaulipas, Mexico." Grant Agreements, 201814J833 (and list all the grant agreement identifiers.)

West, M. and J. Harrison. (1997) *Bayesian Forecasting and Dynamic Models*, 2nd ed. Springer Verlag, New York.

Supplemental material

Appendix 1. Nest counts used in the analysis. These were extracted from the "KR in MX" sheet in MX Beach KR Counts TS.xlsx dated 3/24/2015 and "Sheet 1" in TX Beach KR counts DS.xlsx dated 3/13/2015. Sum3 is the sum of the three beaches monitored since 1991 (Tepehuajes, Rancho Nuevo, and Barra del Tordo), SumMX is the sum of all six Mexican beaches, and SumAll is the sum of all nesting beaches.

Year	La Pesca	Tepehuajes	Rancho Nuevo	Barra del Tordo	Miramar	Altamira	Texas	Sum3	SumMX	SumAll
1991		190	831	157				1178		
1992		276	897	102				1275		
1993		274	857	110				1241		
1994		321	1148	93				1562		
1995		345	1430	155				1930		
1996	20	474	1288	219		79		1981		
1997	39	358	1514	349	3	76	9	2221	2339	2348
1998	40	642	2409	431	15	206	13	3482	3743	3756
1999	57	793	2298	278	16	131	16	3369	3573	3589
2000	144	1622	3778	434	62	130	12	5834	6170	6182
2001	57	708	3742	477	65	246	8	4927	5295	5303
2002	117	989	4012	524	140	490	38	5525	6272	6310
2003	228	1528	5380	696	157	200	19	7604	8189	8208
2004	139	1109	4333	867	187	285	42	6309	6920	6962
2005	245	1610	6947	679	223	311	50	9236	10015	10065
2006	369	2013	7866	1083	402	332	102	10962	12065	12167
2007	150	1624	11268	1213	345	432	128	14105	15032	15160
2008	495	2541	11739	2031	487	589	195	16311	17882	18077
2009	361	1647	16273	2017	431	408	197	19937	21137	21334
2010	202	1221	9840	1313	320	406	141	12374	13302	13443
2011	377	1323	16709	1329	330	502	199	19361	20570	20769
2012	377	1630	16983	1584	612	611	209	20197	21797	22006
2013	236	976	11198	3112	318	547	153	15286	16387	16540
2014	128	1382	7272	2333	427	511	119	10987	12053	12172

Appendix 2. Evaluation of nest count variance and autocorrelation.

Because the response is a count, a preliminary model with a log linear mean and a Poisson response distribution was fit to the data (Equation S1).

$$N_t \sim Poisson(\mu_t) \qquad\qquad\qquad\qquad (S1)$$
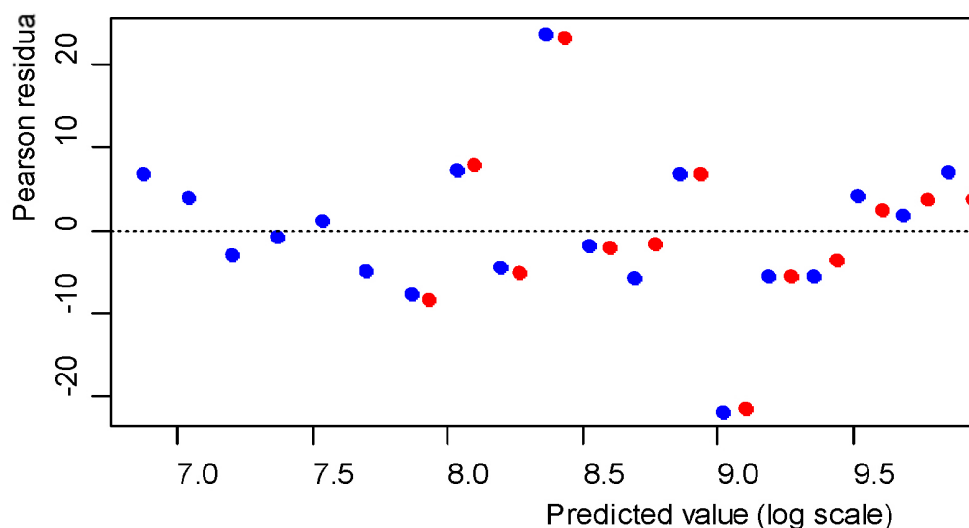
$$\log \mu_t = \beta_0 + \beta_1\, t,$$

where N_t is the nest count in year t and μ_t is the mean nest count for year t.

Inspection of the standardized (Pearson) residuals from this model indicates no concerns with the assumption of constant variance: the spread in the residuals is more or less the same across the range of predicted values (Figure S1).

Many of the standardized residuals from the regression are smaller than -3 or larger than +3, and the standard deviation exceeds 9, considerably larger than the value of 1 that is expected when the observations follow a Poisson distribution. Because of the overdispersion, the generally large nest counts, and the log-linear trend in the mean nest count, subsequent models assume that the log transformed nest count follows a normal response distribution with a standard deviation that is estimated from the data. Similar results are obtained using a negative binomial response distribution, another option to model overdispersed counts (results not shown). The normal distribution for log transformed responses was chosen because modeling correlations is more straightforward when the response is modeled using a normal distribution.

Figure S1. Pearson (standardized) residuals from a log-linear Poisson regression of the annual number of nests for Tepehuajes, Rancho Nuevo, and Barra del Tordo beaches summed (blue) and all nests (red).


The second preliminary model is a linear regression model for log transformed nest counts (Equation S2).

21

$$\log N_t = \beta_0 + \beta_1\, t + \varepsilon_t \qquad\qquad\qquad (S2)$$

$$\varepsilon_t \sim Normal(0,\, \sigma^2)$$

Residuals from this model show no issues with the assumption of constant variance (Figure S2). This result supports the use of log transformed nest count as the response variable. Plots of log-transformed nest counts over time for each beach and beach sum are in the Supplemental Material, Figure S4.

Figure S2. Residuals from a linear regression of the log transformed annual number of nests for Tepehuajes, Rancho Nuevo, and Barra del Tordo beaches summed (blue) and all beaches (red).
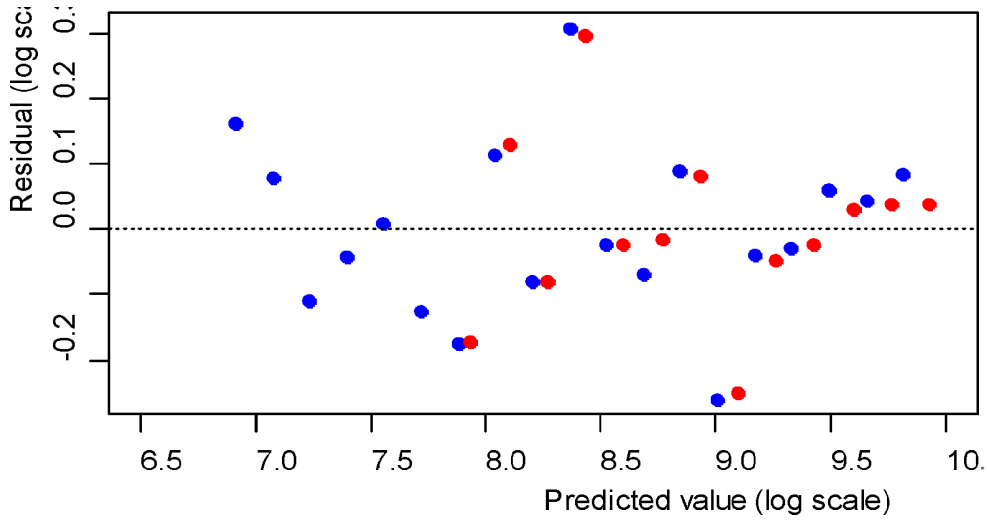
Lack of fit of the linear model was evaluated by fitting a quadratic model (Equation S3) to the nest counts. If the quadratic coefficient is significantly different from 0, there is evidence of lack of fit of the linear model (Equation S2).

$$\log N_t = \beta_0 + \beta_1\, t + \beta_2\, t^2 + \varepsilon_t \qquad\qquad\qquad (S3)$$

$$\varepsilon_t \sim Normal(0,\, \sigma^2)$$

There is no evidence of lack of fit of the linear model. The p-values for the test of $\beta_2 = 0$ are 0.34 for Sum3 and 0.94 for SumAll. The same conclusion is reached for all beaches (Table S1) except Miramar, where the log transformed number of nests increased rapidly from 1997 to 2002, then increased more slowly (Figure S4).

22

Table S1. Diagnostic test results for linear regressions (Equation S2) fit to each beach and beach sum.
LOF p-value is the p-value for lack-of-fit based on the quadratic coefficient, β_2, in Equation (S3). Large p-values indicate no evidence of lack of fit of the linear regression. Lag 1 correlation is the correlation between residuals from the linear regression from adjacent years (e.g. 2002 and 2003). AICc statistics are for three possible models for the correlation between observations over time: uncorrelated, AR(1), and MA(1). The smallest AICc value is underlined and indicates the most appropriate of these three models for each beach or beach sum.

| Beach	LOF	Lag 1	AICc statistics		
	p-value	Correlation	Uncorrelated	AR(1)	MA(1)
sum3	0.34	-0.14	<u>-4.31</u>	-1.60	-1.59
sumAll	0.94	-0.32	<u>2.85</u>	5.56	5.78
sumMx	0.96	-0.33	<u>2.94</u>	5.62	5.86
La Pesca	0.20	-0.44	<u>27.97</u>	29.42	29.63
Tep	0.11	-0.19	<u>19.56</u>	22.12	22.17
Rancho Nuevo	0.10	-0.10	<u>-0.52</u>	2.22	2.22
Barro del Tordo	0.47	0.14	<u>17.56</u>	18.70	19.00
Miramar	0.0002	0.26	34.37	<u>31.42</u>	36.16
Altamira	0.22	-0.16	<u>23.37</u>	26.44	26.44
Texas	0.08	0.00	<u>27.33</u>	30.21	30.34

Because the data are collected over time, we evaluate whether it is necessary to account for autocorrelation among residuals. We do this two ways: using the autocorrelation function (i.e., the correlation between pairs of residuals as a function of the lag between the two observations), and by fitting models with two different correlation structures. These two approaches complement each other. The autocorrelation function provides a visual assessment of correlations at multiple lags but is based on residuals from a model fit assuming no correlation. Model fitting incorporates possible correlation into the model fit but requires specifying the correlation model.

The autocorrelation functions indicate no concerns about correlation between adjacent observations (Figure S3). The most extreme correlations are at lag 4 (observations separated by 4 years, e.g. 2002 and 2006), but both (Sum3 and SumAll) are within the variability expected if observations are uncorrelated and there is no biological basis for such a correlation. The estimated autocorrelations between residuals in adjacent years (e.g. 2001 and 2002) and at lag 2 (e.g. 2001 and 2003) are also within the variability expected from estimates from 19 (Sum3) or 13 (SumAll) uncorrelated observations.

We consider three models to evaluate autocorrelation in the error terms associated with the linear regression model. Two models with correlated errors are the autoregressive order 1, AR(1), model and the moving average order 1, MA(1), model. Under the AR(1) model, all pairs of observations are correlated, but the correlation is most extreme for adjacent observations (e.g. 2002 and 2003). Under the MA(1) model the only non-zero correlation is between adjacent observations. The fit of these

23

models is compared to the fit of the model with no correlation (Equation 2) using the AICc statistic (small-sample-corrected AIC statistic). The model with the smallest AICc value is the most appropriate of the models under consideration. The uncorrelated model has the smallest AICc value for both the Sum3 and the SumAll nest count (Table S1). Similar patterns are seen for all other beaches and beach sums except for Miramar, where the AR(1) model is the most appropriate.

Figure S3. Plot of the autocorrelation between pairs of residuals from the log-linear model (Equation 2) separated by different lags. Blue lines show results for Sum3; red lines show results for SumAll. Lag 1 pairs are those for adjacent years, lag 2 pairs are those 2 years apart, and so forth. The dashed lines indicate approximate critical values for an α = 0.05 test of no autocorrelation.
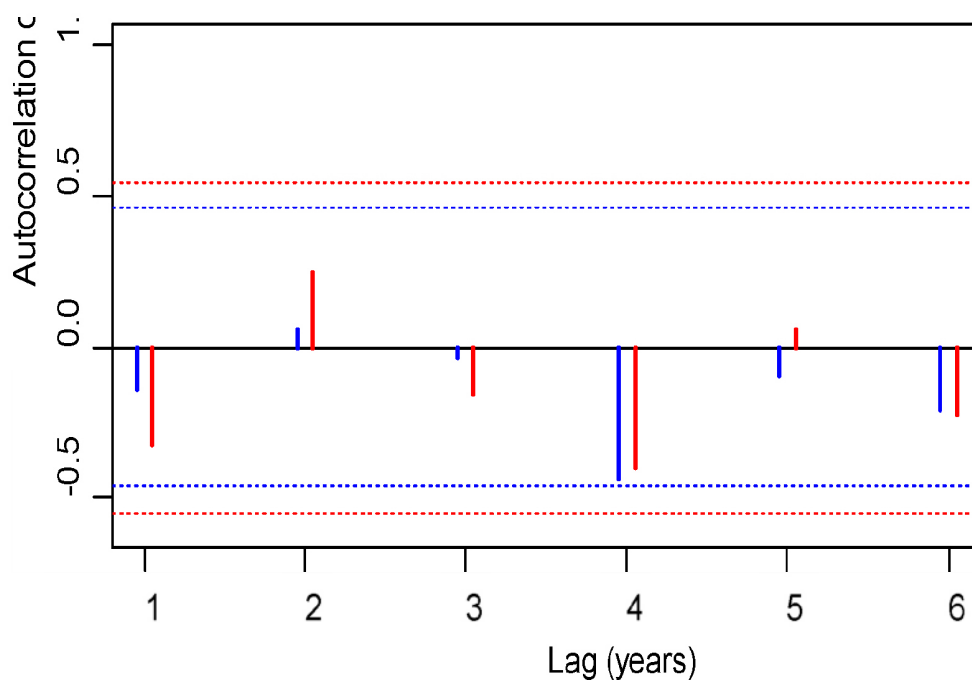
Appendix 3. Estimated regression coefficients from fitting Equation (4) to each individual beach and two beach sums. Results for SumAll are given in Table 3. The estimated regression coefficients for each indicator variable give the difference on a log scale between the projected number of nests and what was observed in that year. These log-scale differences can be expressed as percentages of the expected nest count by exponentiating the estimate. The 95% confidence interval is calculated by exponentiating the 95% confidence interval for the regression estimate.

Beach	Year	Estimate	Std. Error	p value	Percent of Expected	95% Confidence Interval
La.Pesca	2010	-1.072	0.476	0.0437	34%	(13%, 87%)
La.Pesca	2011	-0.66	0.488	0.2013	52%	(20%, 135%)
La.Pesca	2012	-0.872	0.502	0.1078	42%	(16%, 112%)
La.Pesca	2013	-1.553	0.517	0.0110	21%	(8%, 58%)
La.Pesca	2014	-2.377	0.533	0.0008	9%	(3%, 26%)
Tep	2010	-0.878	0.298	0.0090	42%	(23%, 75%)
Tep	2011	-0.933	0.303	0.0067	39%	(22%, 71%)
Tep	2012	-0.859	0.307	0.0124	42%	(23%, 77%)
Tep	2013	-1.507	0.312	0.0002	22%	(12%, 41%)
Tep	2014	-1.294	0.318	0.0008	27%	(15%, 51%)
RN	2010	-0.501	0.165	0.0075	61%	(44%, 84%)
RN	2011	-0.138	0.168	0.4226	87%	(63%, 121%)
RN	2012	-0.288	0.17	0.1089	75%	(54%, 105%)
RN	2013	-0.871	0.173	0.0001	42%	(30%, 59%)
RN	2014	-1.47	0.176	0.0000	23%	(16%, 32%)
BT	2010	-0.503	0.281	0.0911	60%	(35%, 105%)
BT	2011	-0.657	0.285	0.0342	52%	(30%, 91%)
BT	2012	-0.647	0.29	0.0394	52%	(30%, 92%)
BT	2013	-0.137	0.295	0.6479	87%	(49%, 155%)
BT	2014	-0.591	0.3	0.0653	55%	(31%, 100%)
Altamira	2010	-0.457	0.394	0.2684	63%	(29%, 137%)
Altamira	2011	-0.379	0.404	0.3672	68%	(31%, 151%)
Altamira	2012	-0.316	0.416	0.4613	73%	(32%, 165%)
Altamira	2013	-0.561	0.428	0.2145	57%	(25%, 132%)
Altamira	2014	-0.763	0.441	0.1092	47%	(20%, 111%)
Texas	2010	-0.562	0.497	0.2818	57%	(22%, 151%)
Texas	2011	-0.496	0.512	0.3537	61%	(22%, 166%)
Texas	2012	-0.724	0.528	0.1977	48%	(17%, 136%)
Texas	2013	-1.314	0.546	0.0348	27%	(9%, 78%)
Texas	2014	-1.843	0.565	0.0076	16%	(5%, 48%)
sum3	2010	-0.554	0.148	0.0016	57%	(43%, 77%)
sum3	2011	-0.268	0.15	0.0916	76%	(57%, 103%)
sum3	2012	-0.387	0.152	0.0210	68%	(50%, 92%)

Beach	Year	Estimate	Std. Error	p value	Percent of Expected	95% Confidence Interval
sum3	2013	-0.827	0.155	0.0001	44%	(32%, 59%)
sum3	2014	-1.319	0.158	0.0000	27%	(20%, 36%)
sumMx	2010	-0.591	0.164	0.0042	55%	(40%, 76%)
sumMx	2011	-0.32	0.169	0.0843	73%	(52%, 101%)
sumMx	2012	-0.428	0.174	0.0319	65%	(46%, 92%)
sumMx	2013	-0.879	0.18	0.0005	42%	(29%, 59%)
sumMx	2014	-1.352	0.187	0.0000	26%	(18%, 37%)

Appendix 4. Differences in observed and projected nest counts for each beach and beach sums, except for Miramar and SumAll **(in Table 3).** Positive numbers are numbers of "lost" nests (projected – observed). Negative numbers represent more nests than expected.

Beach	Year	Nest Difference Per Year (projected – observed)		Cumulative Nest Difference	
		Median	95% Prediction Interval	Median	95% Prediction Interval
La.Pesca	2010	400	(-10, 1600)	400	(-10, 1600)
La.Pesca	2011	350	(-130, 1910)	820	(30, 2960)
La.Pesca	2012	530	(-90, 2450)	1420	(130, 4640)
La.Pesca	2013	890	(100, 3560)	2390	(470, 7410)
La.Pesca	2014	1290	(250, 4880)	3780	(1030, 10870)
Tep	2010	1730	(330, 4590)	1730	(330, 4590)
Tep	2011	2040	(470, 5160)	3910	(1430, 8310)
Tep	2012	2220	(290, 5910)	6320	(2500, 12300)
Tep	2013	3440	(1290, 7710)	9880	(4740, 18090)
Tep	2014	3670	(1130, 8820)	13630	(6960, 25210)
RN	2010	6500	(1520, 13310)	6500	(1520, 13310)
RN	2011	2420	(-3430, 11210)	9050	(470, 20730)
RN	2012	5580	(-1200, 15230)	15130	(2450, 31650)
RN	2013	15360	(7070, 27640)	30760	(12260, 53690)
RN	2014	24340	(14460, 38680)	55410	(31330, 86020)
BT	2010	870	(-150, 2790)	870	(-150, 2790)
BT	2011	1230	(20, 3540)	2170	(350, 5330)
BT	2012	1430	(60, 3920)	3730	(960, 8130)
BT	2013	450	(-1200, 3920)	4370	(390, 10660)
BT	2014	1900	(-160, 5560)	6350	(1160, 14700)
Miramar	2010	23850	(16390, 35320)	23850	(16390, 35320)
Miramar	2011	28050	(18900, 41540)	52140	(38590, 73020)
Miramar	2012	32860	(21690, 50930)	85230	(64340, 118460)
Miramar	2013	39310	(25840, 60120)	125100	(93730, 170450)
Miramar	2014	46280	(29690, 71850)	171340	(128390, 231950)
Altamira	2010	230	(-160, 1170)	230	(-160, 1170)
Altamira	2011	220	(-230, 1290)	500	(-250, 2090)
Altamira	2012	220	(-300, 1690)	780	(-370, 3250)
Altamira	2013	410	(-200, 2070)	1250	(-350, 4780)
Altamira	2014	570	(-110, 2650)	1880	(-220, 6590)
Texas	2010	110	(-70, 660)	110	(-70, 660)
Texas	2011	120	(-100, 850)	260	(-100, 1280)
Texas	2012	220	(-80, 1310)	520	(-100, 2210)
Texas	2013	410	(10, 1840)	970	(20, 3580)

27

Beach	Year	Nest Difference Per Year (projected – observed)		Cumulative Nest Difference	
		Median	95% Prediction Interval	Median	95% Prediction Interval
Texas	2014	620	(90, 2780)	1660	(250, 5670)
sum3	2010	9150	(3160, 17090)	9150	(3160, 17090)
sum3	2011	6040	(-810, 15530)	15350	(4630, 28050)
sum3	2012	9560	(1070, 20880)	24980	(9130, 44970)
sum3	2013	19640	(9550, 32670)	45050	(24620, 71860)
sum3	2014	29850	(17820, 46520)	75350	(46660, 111550)
sumMx	2010	10870	(3400, 22340)	10870	(3400, 22340)
sumMx	2011	7810	(-1340, 21300)	18910	(5360, 39790)
sumMx	2012	11670	(510, 29740)	30830	(9930, 64050)
sumMx	2013	23240	(9770, 44050)	54620	(23260, 99980)
sumMx	2014	34660	(18060, 60230)	89240	(46290, 149850)

Figure S4. Plots of nest counts over time for all beaches and the SumMX beach sum. Note: the Y axis is on a log scale.
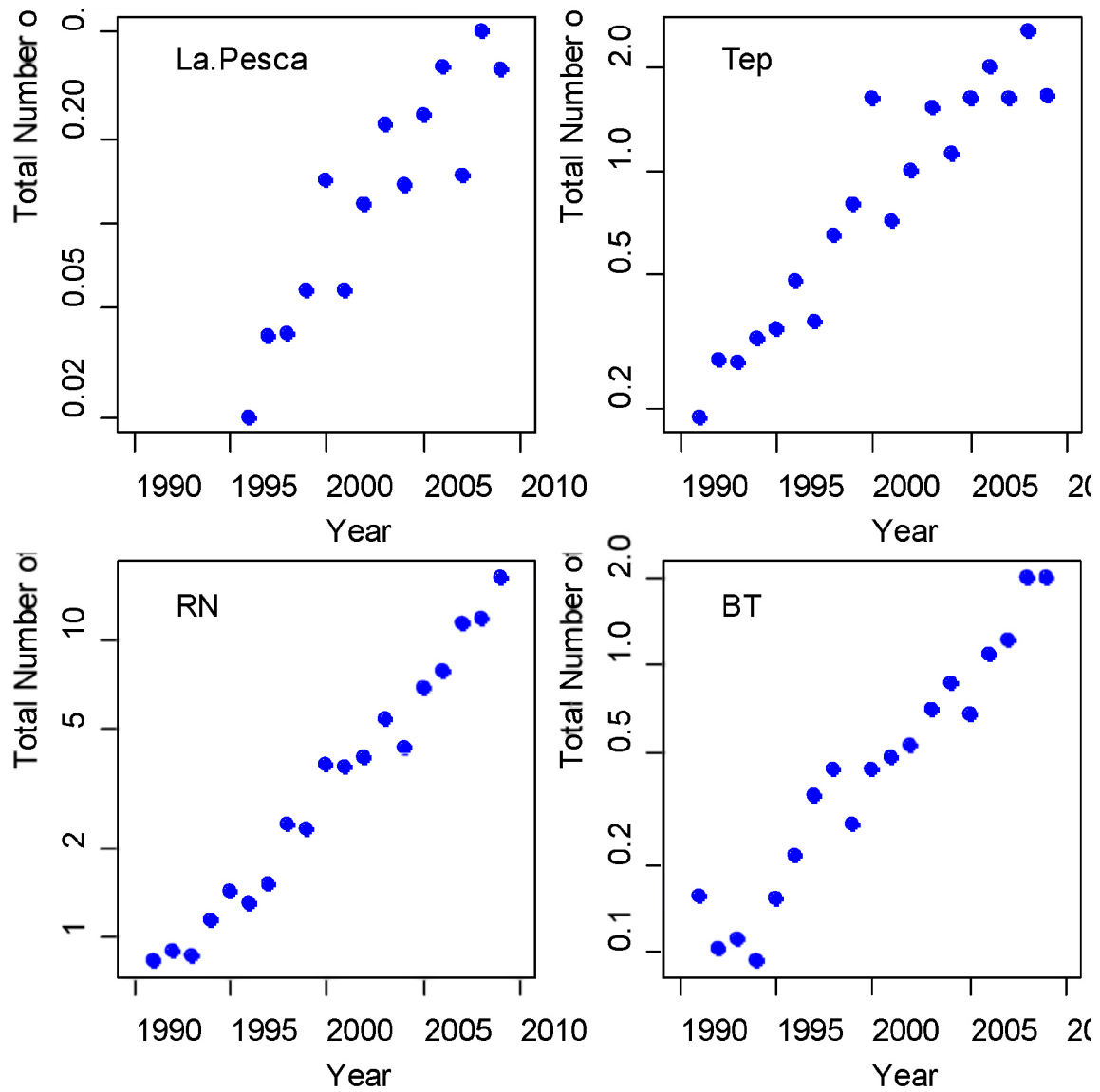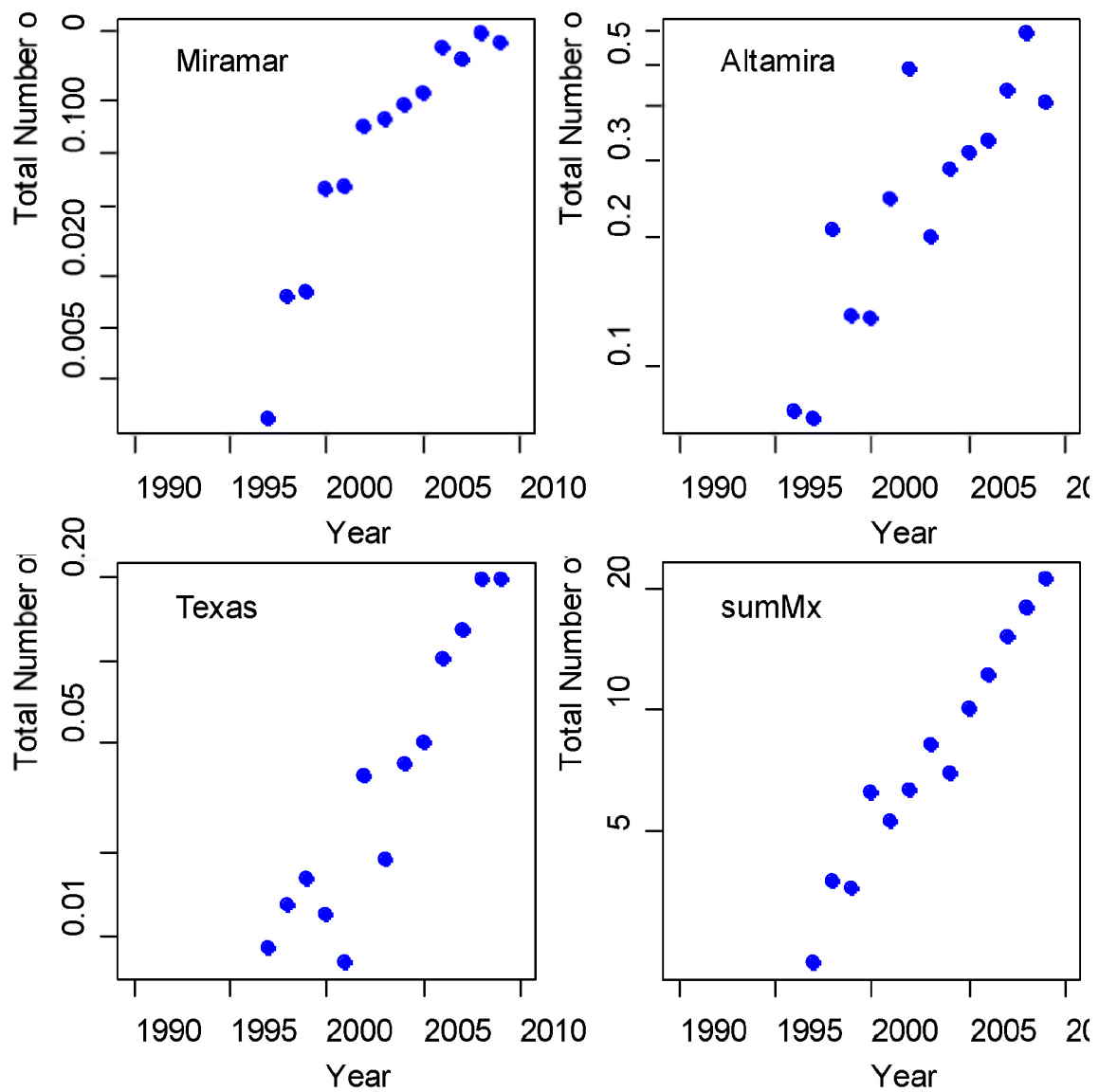
Figure S4 continued.

Figure S5. Plots of the prediction intervals for 2010 through 2014 for each individual beach and the Sum3 and SumMX beach sums. Predictions intervals are computed using the varying trend model (Equations 8-10) for Miramar, because that because showed substantial lack of fit to the linear model and evidence of autocorrelated errors. Prediction intervals for all other beaches and for the two beach sums are computed using the constant trend model.

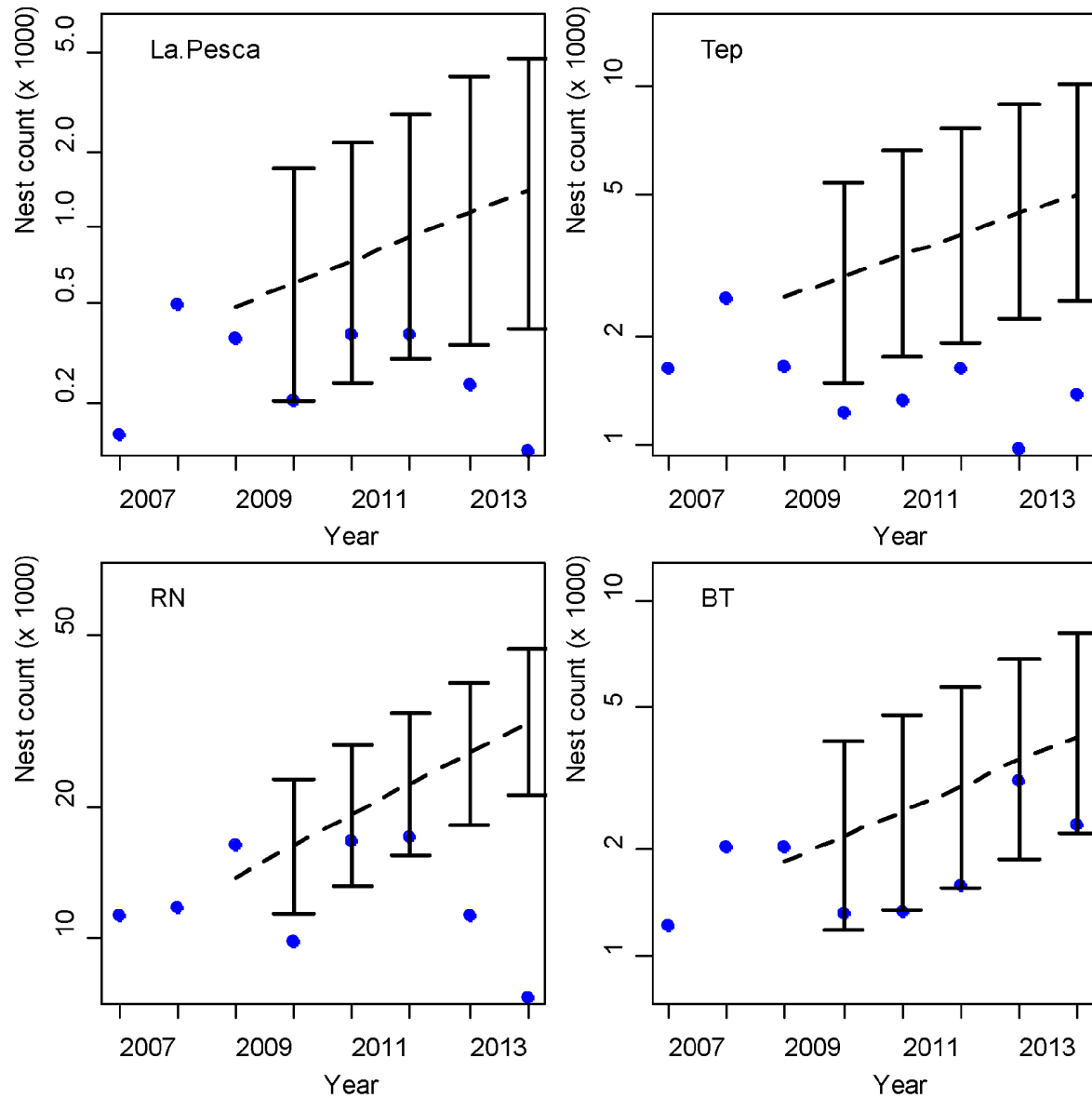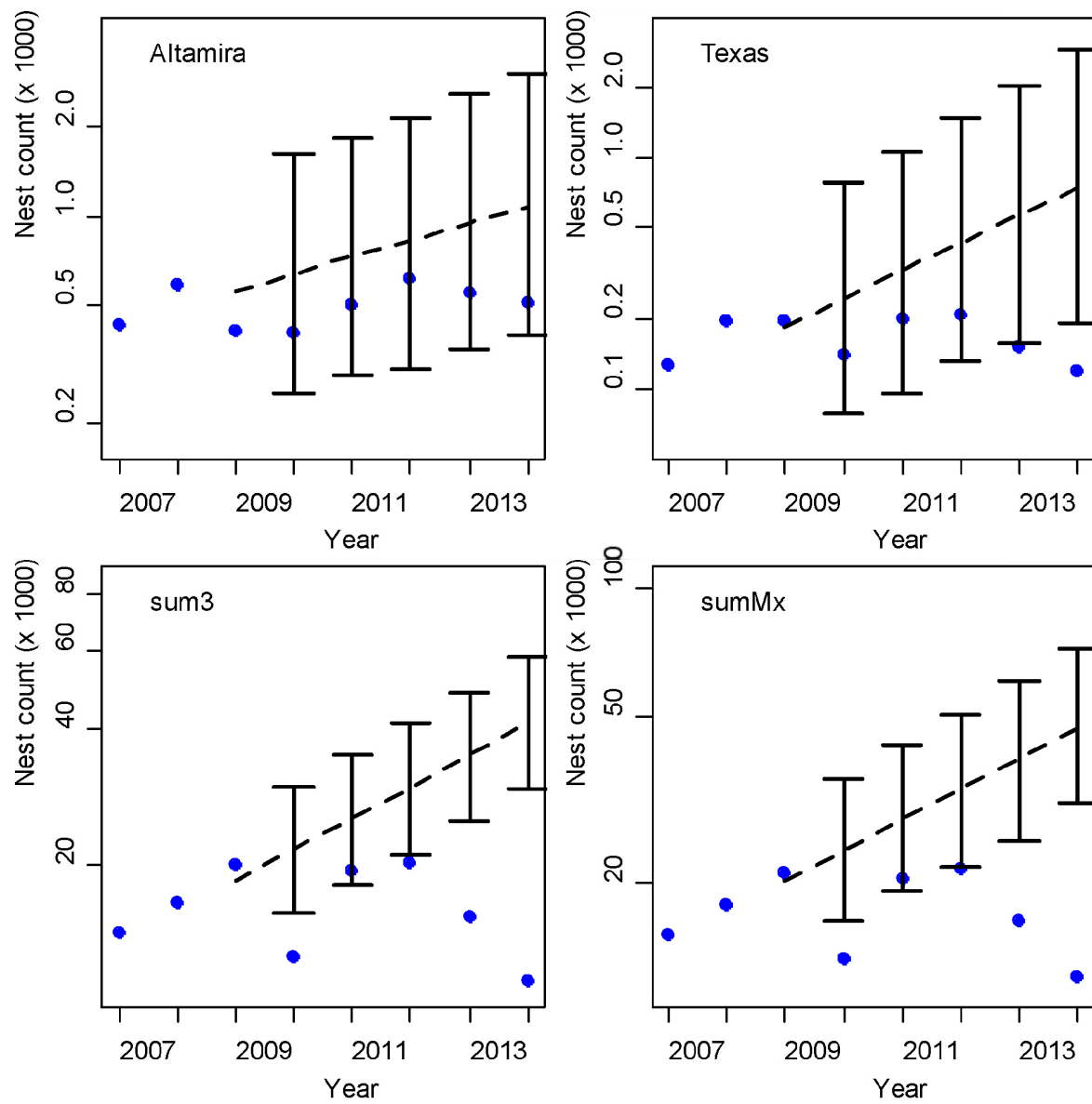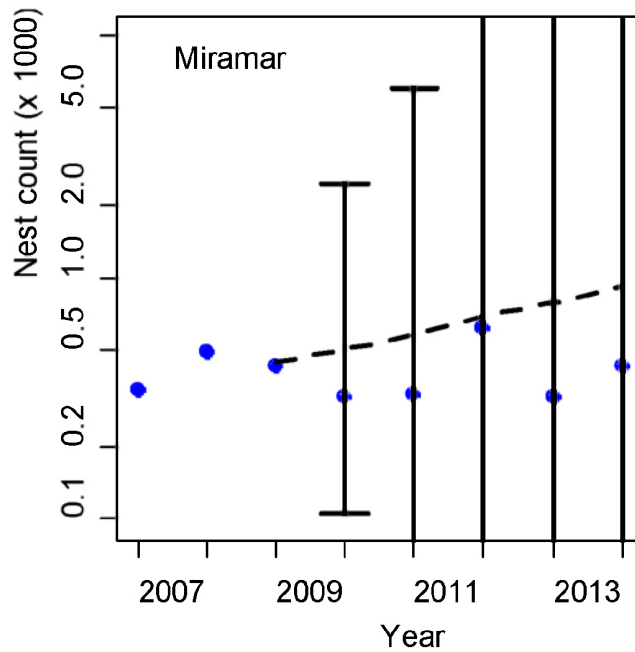**Figure S5 continued.**

32

Figure S5 continued.

Prediction intervals for Miramar are large in 2011 and very large in 2012 and later. Those intervals without upper or lower caps indicate where the interval extends outside the plot boundaries.

Appendix 5: R and BUGS code used for the analyses.

```
# kemp lm.r: fit various regr in two forms:
#   all years with post-oil indicators
#   up through 2009

library(lme4)
library(nlme)

fit.glmod <- fit.lm <- fit.lm2 <- fit.ar <- fit.ma <- fit.ma2 <-
  fit09.lm <- fit09.lm2 <- fit09.ar <- fit09.ma <- fit09.ma2 <-
  as.list(rep(0, 10))

kr$yearc <- kr$Year - 1991
kr$I2010 <- (kr$Year == 2010)+0
kr$I2011 <- (kr$Year == 2011)+0
kr$I2012 <- (kr$Year == 2012)+0
kr$I2013 <- (kr$Year == 2013)+0
kr$I2014 <- (kr$Year == 2014)+0
kr$id <- as.factor(1:24)

kr09$yearc <- kr09$Year - 1991

for (i in 1:10) {
  kr$ncount <- kr[,i+1]
  kr$lncount <- kr[,i+1]

  fit.glmod[[i]] <- glm(ncount ~ yearc + I2010 + I2011 + I2012 + I2013 +
    I2014, data=kr, family=quasipoisson)
  fit.lm[[i]] <- gls(log(ncount) ~ yearc + I2010 + I2011 + I2012 + I2013 +
    I2014, data=kr, na.action=na.omit)
  fit.lm2[[i]] <- gls(log(ncount) ~ yearc + I(yearc^2) + I2010 + I2011 + I2012 + I2013 +
    I2014, data=kr, na.action=na.omit)

  fit.ar[[i]] <- gls(log(ncount) ~ yearc + I2010 + I2011 + I2012 + I2013 +
    I2014, data=kr, corr=corAR1(), na.action=na.omit )
  fit.ma[[i]] <- gls(log(ncount) ~ yearc + I2010 + I2011 + I2012 + I2013 +
    I2014, data=kr, corr=corARMA(p=0, q=1), na.action=na.omit )
  fit.ma2[[i]] <- gls(log(ncount) ~ yearc + I2010 + I2011 + I2012 + I2013 +
    I2014, data=kr, corr=corARMA(p=0, q=2), na.action=na.omit )

  kr09$ncount <- kr09[,i+1]
  kr09$lncount <- log(kr09[,i+1])

  fit09.lm[[i]] <- gls(log(ncount) ~ yearc, data=kr09, na.action=na.omit)
  fit09.lm2[[i]] <- gls(log(ncount) ~ yearc +I(yearc^2), data=kr09, na.action=na.omit)
  fit09.ar[[i]] <- gls(log(ncount) ~ yearc, data=kr09, corr=corAR1(), na.action=na.omit )
  fit09.ma[[i]] <- gls(log(ncount) ~ yearc, data=kr09, corr=corARMA(p=0, q=1),
```

34

```
    na.action=na.omit )
  fit09.ma2[[i]] <- gls(log(ncount) ~ yearc, data=kr09, corr=corARMA(p=0, q=2),
    na.action=na.omit )
  }

lof.all <- sapply(fit09.lm2, function(x){summary(x)$tTable[3,4]})
lag1.all <- sapply(fit09.lm, function(x) {
  acf(resid(x))$acf[2,1,1]})

aic.all <- rbind( lm = sapply(fit.lm, AIC), ar1 = sapply(fit.ar, AIC),
  ma1 = sapply(fit.ma, AIC), ma2=sapply(fit.ma2, AIC) )
dimnames(aic.all)[[2]] <- beachnames

aic.all09 <- rbind( lm = sapply(fit09.lm, AIC), ar1 = sapply(fit09.ar, AIC),
  ma1 = sapply(fit09.ma, AIC), ma2 = sapply(fit09.ma2, AIC) )
dimnames(aic.all09)[[2]] <- beachnames

write.csv(t(rbind(lof=lof.all, lag1=lag1.all, aic.all09[-4,])), file='diag.csv')
```

```
# kemp.r: Read data files,
#  Fit all beaches individually, only data through 2009
#  fits process error and varying slope models

mx <- read.csv('MX beaches.csv', as.is=T)
tx <- read.csv('TX Beach KR counts DS.csv', as.is=T)

kr <- cbind(mx, Texas=c(rep(NA, 6), tx$nests),
  sum3 = mx$Tep + mx$RN + mx$BT, sumMx = apply(mx[,2:7],1, sum)
  )

kr$sumAll <- kr$sumMx + kr$Texas

kr09 <- kr[kr$Year <= 2009, ]

beachnames <- names(kr09)[-1]

kemp.param <- c('sigma', 'b0', 'l0', 'my')
kempb.param <- c('sigma', 'b', 'b0', 'l0', 'my')
kempKM.param <- c('sigma', 'b', 'b0', 'l0', 'my', 'z')

kemp.sim.inits <- function(n) {
  # initial values when precision given 1/Unif^2, as per Gelman
  # for kemp2grp.txt model (no uncertain shift in 2010)

 list( b0 = rnorm(1), sb=runif(1), sl=runif(1), sy = runif(1) )
}

kemp.inits <- list(
  kemp.sim.inits(n), kemp.sim.inits(n), kemp.sim.inits(n) )

kempKM.sim.inits <- function(n) {
  # initial values when precision given 1/Unif^2, as per Gelman
  # for kemp2grp.txt model (no uncertain shift in 2010)

 list( b0 = rnorm(1), sb=runif(1), sl=runif(1), sy = runif(1),
    Model = 1+rbinom(1,1, 0.5) )
}

kempKM.inits <- list(
  kempKM.sim.inits(n), kempKM.sim.inits(n), kempKM.sim.inits(n) )


neach <- 30000

nbeach <- dim(kr09)[2]-1  # number of beaches in kr09

fit09.beach <- fit09.beachb <- as.list(rep(0, nbeach))
```

36

```
for (i in 1:nbeach) {
  nests <- kr09[,i+1]  # number of nests on beach i
  if (any(is.na(nests))) {
   nests <- nests[-(1:sum(is.na(nests)))]
   }
  # remove NA's at beginning of sequence

  kemp.data <- list(n=length(nests), y=log(nests), pi.model=c(0.5,0.5) )

# constant slope model
  fit09.beach[[i]] <- bugs(data = kemp.data, inits = kemp.inits,
         parameters.to.save = kemp.param,
      model.file = "kemp.txt",
         n.chains = 3,
         n.iter = 10000+neach,
         n.burnin = 10000,
         n.sims=neach/10,
         debug = TRUE,
         bugs.directory = "C:/Program Files/WinBUGS14/",
#    program = "WinBUGS"
   program = "OpenBUGS"
             )

# b follows random walk

  fit09.beachb[[i]] <- bugs(data = kemp.data, inits = kemp.inits,
         parameters.to.save = kempb.param,
         model.file = "kempb.txt",
         n.chains = 3,
         n.iter = 10000+neach,
         n.burnin = 10000,
         n.sims=neach/10,
         debug = TRUE,
         bugs.directory = "C:/Program Files/WinBUGS14/",
#    program = "WinBUGS"
   program = "OpenBUGS"
             )

}

all.DIC09 <- matrix(NA, nrow=2, ncol=nbeach)
all.DIC09[1,] <- sapply(fit09.beach, function(x){x$DIC})
all.DIC09[2,] <- sapply(fit09.beachb, function(x){x$DIC})
dimnames(all.DIC09) <- list( c('constant', 'varies'), beachnames)

all.pd09 <- matrix(NA, nrow=4, ncol=nbeach)
all.pd09[1,] <- sapply(fit09.beach, function(x){x$pD})
```

```
all.pd09[2,] <- sapply(fit09.beachb, function(x){x$pD})
all.pd09[3,] <- sapply(fit09.beach, function(x){temp <- x$sims.list$deviance;
  mean(temp) + var(temp)/2 })
all.pd09[4,] <- sapply(fit09.beachb, function(x){temp <- x$sims.list$deviance;
  mean(temp) + var(temp)/2 })

dimnames(all.pd09) <- list( c('constant: BUGS', 'varies: BUGS',
  'constant: Gelman', 'varies: Gelman'), beachnames)
```

```
# kemp2.r: Fit all beaches individually, only data through 2009
# only fitting log linear regression model in state space form

# relies on data sets setup by kemp.r

kemp2.param <- c('sigma', 'b0', 'l')

kemp2.sim.inits <- function(n) {
  # initial values when precision given 1/Unif^2, as per Gelman
  # for kemp2.txt model

 list( b0 = rnorm(1), sy = runif(1) )
}

kemp2.inits <- list(
  kemp2.sim.inits(n), kemp.sim.inits(n), kemp.sim.inits(n) )

neach <- 30000

fit09.beach2 <- as.list(rep(0, nbeach))

for (i in 1:nbeach) {
  nests <- kr09[,i+1]  # number of nests on beach i
  if (any(is.na(nests))) {
    nests <- nests[-(1:sum(is.na(nests)))]
    }
  # remove NA's at beginning of sequence

  kemp.data <- list(n=length(nests), y=log(nests))

# log linear regression model
  fit09.beach2[[i]] <- bugs(data = kemp.data, inits = kemp2.inits,
         parameters.to.save = kemp2.param,
     model.file = "kemp2.txt",
         n.chains = 3,
         n.iter = 10000+neach,
         n.burnin = 10000,
         n.sims=neach/10,
         debug = TRUE,
         bugs.directory = "C:/Program Files/WinBUGS14/",
#    program = "WinBUGS"
   program = "OpenBUGS"
             )

}
```

39

```
# kemp.txt  BUGS code for process error model

model
{
l[1] <- l0
my[1] ~ dnorm(l[1], taul)

for (i in 2:n) {

  l[i] <- my[i-1] + b0
  my[i] ~ dnorm(l[i], taul)
  }

for (i in 1:n) {
  y[i] ~ dnorm(my[i], tauy)
  }

# priors
l0 ~ dnorm(0, 0.001)
b0 ~ dnorm(0, 0.001)

taul <- pow(sl, -2)
sl ~ dunif(0,10)

tauy <- pow(sy, -2)
sy ~ dunif(0,10)

sigma[1] <- 0
sigma[2] <- 1/sqrt(taul)
sigma[3] <- 1/sqrt(tauy)

}
```

40

kemp2.txt BUGS code for constant slope model

```
model
{
l[1] <- l0

for (i in 2:n) {

  l[i] <- l[i-1] + b0
  }

for (i in 1:n) {
  y[i] ~ dnorm(l[i], tauy)
  }

# priors
l0 ~ dnorm(0, 0.001)
b0 ~ dnorm(0, 0.001)

tauy <- pow(sy, -2)
sy ~ dunif(0,100)

sigma <- 1/sqrt(tauy)

}
```

```
# kempb.txt BUGS code for varying slope model

model
{
b[1] <- b0
l[1] <- l0
my[1] ~ dnorm(l[1], taul)

for (i in 2:n) {
  mb[i] <- b[i-1]
  b[i] ~ dnorm(mb[i], taub)

  l[i] <- my[i-1] + b[i]
  my[i] ~ dnorm(l[i], taul)
  }

for (i in 1:n) {
  y[i] ~ dnorm(my[i], tauy)
  }

# priors
l0 ~ dnorm(0, 0.001)
b0 ~ dnorm(0, 0.001)

taub <- pow(sb, -2)
sb ~ dunif(0,10)

taul <- pow(sl, -2)
sl ~ dunif(0,10)

tauy <- pow(sy, -2)
sy ~ dunif(0,10)

sigma[1] <- 1/sqrt(taub)
sigma[2] <- 1/sqrt(taul)
sigma[3] <- 1/sqrt(tauy)


}
```